Student Name : Sankar Samiksha

Group : SS5

Date : 5 April 2022

LAB 4: ANALZING NETWORK DATA LOG

You are provided with the data file, in .csv format, in the working directory. Write the program to extract the following informations.

EXERCISE 4A: TOP TALKERS AND LISTENERS

One of the most commonly used function in analyzing data log is finding out the IP address of the hosts that send out large amount of packet and hosts that receive large number of packets, usually know as TOP TALKERS and LISTENERS. Based on the IP address we can obtained the organization who owns the IP address.

List the TOP 5 TALKERS

Rank	IP address	# of packets	Organisation
1	193.62.192.8	3041	European Bioinformatics, England
2	155.69.160.32	2975	NTU, Singapore
3	130.14.250.11	2604	National Library of Medicine, United States
4	14.139.196.58	2452	Indian Institute of Technology (IIT) Guwahati, India
5	140.112.8.139	2056	Taiwan Academic Network, Taiwan

TOP 5 LISTENERS

Rank	IP address	# of packets	Organisation
1	103.37.198.100	3841	A*STAR, Singapore
2	137.132.228.15	3715	NUS, Singapore
3	202.21.159.244	2446	Republic Polytechnic, Singapore
4	192.101.107.153	2368	Pacific Northwest National Laboratory, United States
5	103.21.126.2	2056	Powai, India

EXERCISE 4B: TRANSPORT PROTOCOL

Using the IP protocol type attribute, determine the percentage of TCP and UDP protocol

	Header value	Transport layer protocol	# of packets
1	6	TCP	56064
2	17	UDP	9462
3	50	ESP	1698

Total Number of Packets in the entire CSV: 69370
Percentage of UDP: 13.639901974917112
Percentage of TCP: 80.81879775118928

EXERCISE 4C: APPLICATIONS PROTOCOL

Using the Destination IP port number determine the most frequently used application protocol. (For finding the service given the port number https://www.adminsub.net/tcp-udp-port-finder/)

Rank	Destination IP port number	# of packets	Service
1	443	13423	https
2	80	2647	http
3	52866	2068	Dynamic and/or private ports (UDP) Xsan. Xsan Filesystem Access(TCP)
4	45512	1356	Unassigned
5	56152	1341	Dynamic and/or private ports Xsan. Xsan Filesystem Access(TCP)

EXERCISE 4D: TRAFFIC

The traffic intensity is an important parameter that a network engineer needs to monitor closely to determine if there is congestion. You would use the IP packet size to calculate the estimated total traffic over the monitored period of 15 seconds. (Assume the sampling rate is 1 in 2048)

IP packet size corresponds to the IP_Size column name

Total Traffic(MB)	132664.98 MB (to 2 d.p)

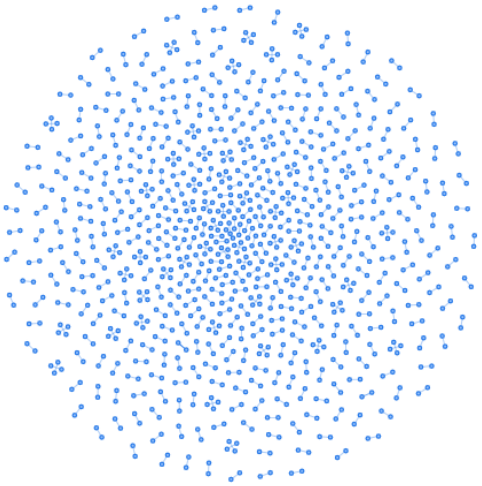
EXERCISE 4E: ADDITIONAL ANALYSIS

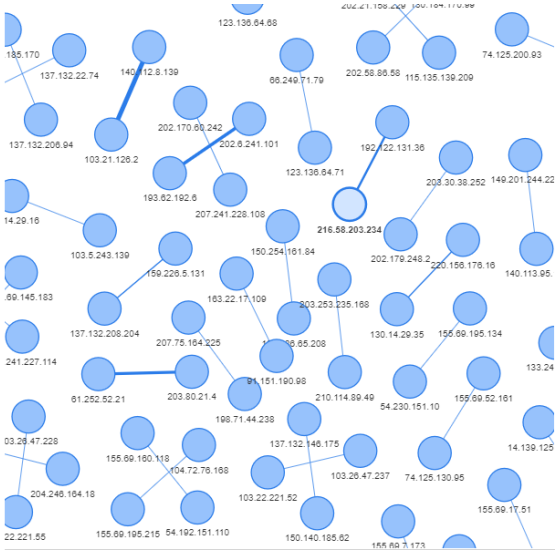
Please append ONE page to provide additional analysis of the data and the insight it provides.
Examples include:
Top 5 communication pairs;
Visualization of communications between different IP hosts;
etc.
Please limit your results within one page (and any additional results that fall beyond one page limit will not be assessed).

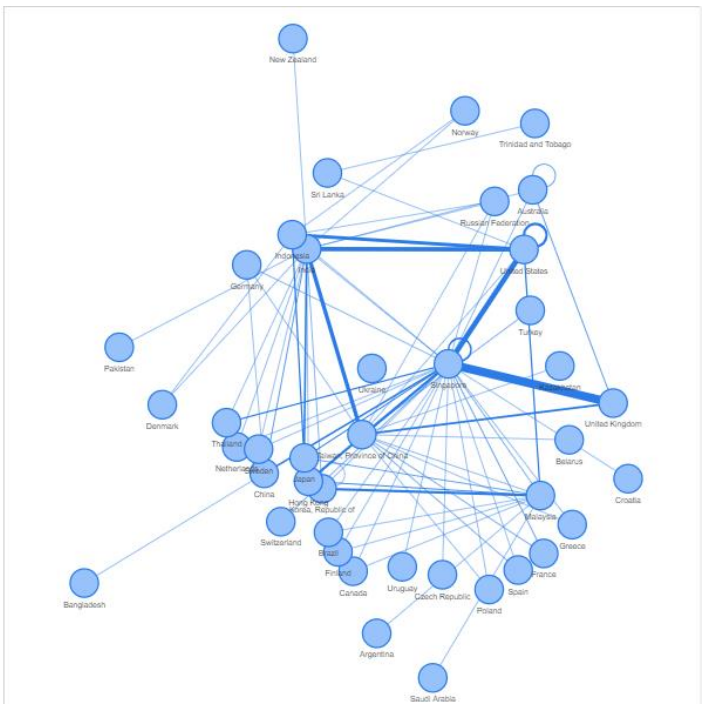
The top 5 communication pairs

IP Address 1	IP Address 2	Count
137.132.228.15	193.62.192.8	4951
130.14.250.11	103.37.198.100	2842
14.139.196.58	192.101.107.153	2368
103.21.126.2	140.112.8.139	2056
167.205.52.8	140.90.101.61	1752

Using the table of communication pairs the network graph below was plotted. The graph is very sparse and most nodes are only connected to one other node. The image in the right is a zoomed in version. The thickness of the line indicates the count or the amount the two ip addresses have interacted with each other.







Later, the ip addresses were grouped accordingly the country of origin. Below shows the graph:
A hollow circle indicates that the two ip addresses interacting with each other come from the same country. Similar to earlier, the thicker the line, the more the interaction. Here we can see that Singapore interacts most with United Kingdom followed by United States.

EXERCISE 4F: SOFTWARE CODE

Please also submit your code to the NTULearn lab site.